



there may be gene-specific exceptions to this rule). RNAi acts systemically — injection into one tissue inhibits gene function in cells throughout the animal. The results of a variety of experiments, in *C. elegans* and other organisms, indicate that RNAi acts to destabilize cellular RNA after RNA processing.

5           The potency of RNAi inspired Timmons and Fire (1998 Nature **395**: 854) to do a simple experiment that produced an astonishing result. They fed to nematodes bacteria that had been engineered to express double-stranded RNA corresponding to the *C. elegans* *unc-22* gene. Amazingly, these nematodes developed a phenotype similar to that of *unc-22* mutants that was dependent on their food source. The ability to conditionally expose large  
10 numbers of nematodes to gene-specific double-stranded RNA formed the basis for a very powerful screen to select for RNAi-defective *C. elegans* mutants and then to identify the corresponding genes.

          Double-stranded RNAs (dsRNAs) can provoke gene silencing in numerous in vivo contexts including *Drosophila*, *Caenorhabditis elegans*, planaria, hydra, trypanosomes,  
15 fungi and plants. However, the ability to recapitulate this phenomenon in higher eukaryotes, particularly mammalian cells, has not been accomplished in the art. Nor has the prior art demonstrated that this phenomena can be observed in cultured eukaryotes cells.

### Summary of the Invention

20           One aspect of the present invention provides a method for attenuating expression of a target gene in cultured cells, comprising introducing double stranded RNA (dsRNA) into the cells in an amount sufficient to attenuate expression of the target gene, wherein the dsRNA comprises a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of the target gene.

25           Another aspect of the present invention provides a method for attenuating expression of a target gene in a mammalian cell, comprising

- (i) activating one or both of a Dicer activity or an Argonaut activity in the cell, and
- (ii) introducing into the cell a double stranded RNA (dsRNA) in an amount  
30 sufficient to attenuate expression of the target gene, wherein the dsRNA comprises a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of the target gene.

In certain embodiments, the cell is suspended in culture; while in other embodiments the cell is in a whole animal, such as a non-human mammal.





- (i) identifying, by the subject assay, a target gene which provides a phenotypically desirable response when inhibited by RNAi;
- (ii) (optionally) conducting therapeutic profiling of the target gene for efficacy and toxicity in animals; and
- 5 (iii). licensing, to a third party, the rights for further drug development of inhibitors of the target gene.

Another aspect of the invention provides a method for inhibiting RNAi by inhibiting the expression or activity of an RNAi enzyme. Thus, the subject method may include inhibiting the acitivity of Dicer and/or the 22-mer RNA.

10 Still another aspect relates to the a method for altering the specificity of an RNAi by modifying the sequence of the RNA component of the RNAi enzyme.

Another aspect of the invention relates to purified or semi-purified preparations of the RNAi enzyme or components thereof. In certain embodiments, the preparations are used for identifying compounds, especially small organic molecules, which inhibit or  
15 potentiate the RNAi activity. Small molecule inhibitors, for example, can be used to inhibit dsRNA responses in cells which are purposefully being transfected with a virus which produces double stranded RNA.

The dsRNA construct may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the  
20 nucleoside. The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The dsRNA construct may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses of double-stranded material may yield more effective inhibition. Inhibition is sequence-specific in that  
25 nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. dsRNA constructs containing a nucleotide sequences identical to a portion of the target gene is preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may optimized by alignment  
30 algorithms known in the art and calculating the percent difference between the nucleotide sequences. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript.

Yet another aspect of the invention pertains to transgenic non-human mammals  
35 which include a transgene encoding a dsRNA construct, preferably which is stably integrated into the genome of cells in which it occurs. The animals can be derived by

oocyte microinjection, for example, in which case all of the nucleated cells of the animal will include the transgene, or can be derived using embryonic stem (ES) cells which have been transfected with the transgene, in which case the animal is a chimera and only a portion of its nucleated cells will include the transgene. In certain instances, the sequence-independent dsRNA response, e.g., the PKR response, is also inhibited in those cells including the transgene.

In still other embodiments, dsRNA itself can be introduced into an ES cell in order to effect gene silencing, and that phenotype will be carried for at least several rounds of division, e.g., into the progeny of that cell.

### **Brief Description of the Drawings**

Figure 1: RNAi in S2 cells. a, *Drosophila* S2 cells were transfected with a plasmid that directs *lacZ* expression from the copia promoter in combination with dsRNAs corresponding to either human CD8 or *lacZ*, or with no dsRNA, as indicated. b, S2 cells were co-transfected with a plasmid that directs expression of a GFP-US9 fusion protein (12) and dsRNAs of either *lacZ* or *cyclin E*, as indicated. Upper panels show FACS profiles of the bulk population. Lower panels show FACS profiles from GFP-positive cells. c, Total RNA was extracted from cells transfected with *lacZ*, *cyclin E*, *fizzy* or *cyclin A* dsRNAs, as indicated. Northern blots were hybridized with sequences not present in the transfected dsRNAs.

Figure 2: RNAi *in vitro*. a, Transcripts corresponding to either the first 600 nucleotides of *Drosophila cyclin E* (E600) or the first 800 nucleotides of *lacZ* (Z800) were incubated in lysates derived from cells that had been transfected with either *lacZ* or *cyclin E* (*cycE*) dsRNAs, as indicated. Time points were 0, 10, 20, 30, 40 and 60 min for *cyclin E* and 0, 10, 20, 30 and 60 min for *lacZ*. b, Transcripts were incubated in an extract of S2 cells that had been transfected with *cyclin E* dsRNA (cross-hatched box, below). Transcripts corresponded to the first 800 nucleotides of *lacZ* or the first 600, 300, 220 or 100 nucleotides of *cyclin E*, as indicated. Eout is a transcript derived from the portion of the *cyclin E* cDNA not contained within the transfected dsRNA. E-ds is identical to the dsRNA that had been transfected into S2 cells. Time points were 0 and 30 min. c, Synthetic transcripts complementary to the complete *cyclin E* cDNA (Eas) or the final 600 nucleotides (Eas600) or 300 nucleotides (Eas300) were incubated in extract for 0 or 30 min.

Figure 3: Substrate requirements of the RISC. Extracts were prepared from cells transfected with *cyclin E* dsRNA. Aliquots were incubated for 30 min at 30 °C before the addition of either the *cyclin E* (E600) or *lacZ* (Z800) substrate. Individual 20- $\mu$ l aliquots,

as indicated, were pre-incubated with 1 mM CaCl<sub>2</sub> and 5 mM EGTA, 1 mM CaCl<sub>2</sub>, 5 mM EGTA and 60 U of micrococcal nuclease, 1 mM CaCl<sub>2</sub> and 60 U of micrococcal nuclease or 10 U of DNase I (Promega) and 5 mM EGTA. After the 30-min pre-incubation, EGTA was added to those samples that lacked it. Yeast tRNA (1 µg) was added to all samples.

5 Time points were at 0 and 30 min.

Figure 4: The RISC contains a potential guide RNA. a, Northern blots of RNA from either a crude lysate or the S100 fraction (containing the soluble nuclease activity, see Methods) were hybridized to a riboprobe derived from the sense strand of the *cyclin E* mRNA. b, Soluble *cyclin-E*-specific nuclease activity was fractionated as described in  
10 Methods. Fractions from the anion-exchange resin were incubated with the lacZ, control substrate (upper panel) or the cyclin E substrate (centre panel). Lower panel, RNA from each fraction was analysed by northern blotting with a uniformly labelled transcript derived from sense strand of the *cyclin E* cDNA. DNA oligonucleotides were used as size markers.

Figure 5: Generation of 22mers and degradation of mRNA are carried out by distinct enzymatic complexes. A. Extracts prepared either from 0-12 hour *Drosophila* embryos or *Drosophila* S2 cells (see Methods) were incubated 0, 15, 30, or 60 minutes (left to right) with a uniformly-labeled double-stranded RNA corresponding to the first 500 nucleotides of the *Drosophila cyclin E* coding region. M indicates a marker prepared by *in vitro* transcription of a synthetic template. The template was designed to yield a 22 nucleotide transcript. The doublet most probably results from improper initiation at the +1 position. B. Whole-cell extracts were prepared from S2 cells that had been transfected with a dsRNA corresponding to the first 500 nt. of the luciferase coding region. S10 extracts were spun at 30,000xg for 20 minutes which represents our standard RISC  
20 extract<sup>6</sup>. S100 extracts were prepared by further centrifugation of S10 extracts for 60 minutes at 100,000xg. Assays for mRNA degradation were carried out as described previously<sup>6</sup> for 0,30 or 60 minutes (left to right in each set) with either a single-stranded luciferase mRNA or a single-stranded cyclin E mRNA, as indicated. C. S10 or S100  
25 extracts were incubated with cyclin E dsRNAs for 0, 60 or 120 minutes (L to R).

Figure 6: Production of 22mers by recombinant CG4792/Dicer. A. *Drosophila* S2 cells were transfected with plasmids that direct the expression of T7-epitope tagged versions of Drosha, CG4792/Dicer-1 and Homeless. Tagged proteins were purified from cell lysates by immunoprecipitation and were incubated with *cyclin E* dsRNA. For comparison, reactions were also performed in *Drosophila* embryo and S2 cell extracts. As  
35 a negative control, immunoprecipitates were prepared from cells transfected with a β-galactosidase expression vector. Pairs of lanes show reactions performed for 0 or 60 minutes. The synthetic marker (M) is as described in the legend to Figure 1. B.

Diagrammatic representations of the domain structures of CG4792/Dicer-1, Drosha and Homeless are shown. C. Immunoprecipitates were prepared from detergent lysates of S2 cells using an antiserum raised against the C-terminal 8 amino acids of Drosophila Dicer-1 (CG4792). As controls, similar preparations were made with a pre-immune serum and with an immune serum that had been pre-incubated with an excess of antigenic peptide. Cleavage reactions in which each of these precipitates was incubated with an ~500 nt. fragment of Drosophila cyclin E are shown. For comparison, an incubation of the substrate in Drosophila embryo extract was electrophoresed in parallel. D. Dicer immunoprecipitates were incubated with dsRNA substrates in the presence or absence of ATP. For comparison, the same substrate was incubated with S2 extracts that either contained added ATP or that were depleted of ATP using glucose and hexokinase (see methods). E. Drosophila S2 cells were transfected with uniformly, <sup>32</sup>P-labelled dsRNA corresponding to the first 500 nt. of GFP. RISC complex was affinity purified using a histidine-tagged version of D.m. Ago-2, a recently identified component of the RISC complex (Hammond et al., in prep). RISC was isolated either under conditions in which it remains ribosome associated (ls, low salt) or under conditions that extract it from the ribosome in a soluble form (hs, high salt)<sup>6</sup>. For comparison, the spectrum of labelled RNAs in the total lysate is shown. F. Guide RNAs produced by incubation of dsRNA with a Dicer immunoprecipitate are compared to guide RNAs present in a affinity-purified RISC complex. These precisely comigrate on a gel that has single-nucleotide resolution. The lane labelled control is an affinity selection for RISC from cell that had been transfected with labeled dsRNA but not with the epitope-tagged D.m. Ago-2.

Figure 7: Dicer participates in RNAi. A. Drosophila S2 cells were transfected with dsRNAs corresponding to the two Drosophila Dicers (CG4792 and CG6493) or with a control dsRNA corresponding to murine caspase 9. Cytoplasmic extracts of these cells were tested for Dicer activity. Transfection with Dicer dsRNA reduced activity in lysates by 7.4-fold. B. The Dicer-1 antiserum (CG4792) was used to prepare immunoprecipitates from S2 cells that had been treated as described above. Dicer dsRNA reduced the activity of Dicer-1 in this assay by 6.2-fold. C. Cells that had been transfected two days previously with either mouse caspase 9 dsRNA or with Dicer dsRNA were cotransfected with a GFP expression plasmid and either control, luciferase dsRNA or GFP dsRNA. Three independent experiments were quantified by FACS. A comparison of the relative percentage of GFP-positive cells is shown for control (GFP plasmid plus luciferase dsRNA) or silenced (GFP plasmid plus GFP dsRNA) populations in cells that had previously been transfected with either control (caspase 9) or Dicer dsRNAs.

Figure 8: Dicer is an evolutionarily conserved ribonuclease. A. A model for production of 22mers by Dicer. Based upon the proposed mechanism of action of Ribonuclease III, we propose that Dicer acts on its substrate as a dimer. The positioning





Figure 21: Dicer requires ATP.

Figure 22: Dicer produces RNAs that are the same size as RNAs present in RISC.

Figure 23: Human dicer homolog when expressed and immunoprecipitated has 22mer generating activity.

5 Figure 24: Sequence of dm argonaute 2. Peptides identified by microsequencing are shown in underline.

Figure 25: Molecular characterization of dm argonaute 2. The presence of an intron in coding sequence was determined by northern blotting using intron probe. This results in a different 5' reading frame than that published genome sequence. Number of polyglutamine repeats was determined by genomic PCR.

Figure 26: Dicer activity can be created in human cells by expression of human dicer gene. Host cell was 293. Crude extracts had dicer activity, while activity was absent from untransfected cells. Activity is not dissimilar to that seen in drosophila embryo extracts..

15 Figure 27: An ~500 nt. fragment of the gene that is to be silenced (X) is inserted into the modified vector as a stable direct repeat using standard cloning procedures. Treatment with commercially available cre recombinase reverses sequences within the loxP sites (L) to create an inverted repeat. This can be stably maintained and amplified in an sbc mutant bacterial strain (DL759). Transcription in vivo from the promoter of choice (P) yields a hairpin RNA that causes silencing. A zeocin resistance marker is included to insure maintenance of the direct and inverted repeat structures; however this is non-essential in vivo and could be removed by pre-mRNA splicing if desired. Smith, N. A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* 407, 319-20 (2000).

25 Figure 28: HeLa, Chinese hamster ovary, and P19 (pluripotent, mouse embryonic carcinoma) cell lines transfected with plasmids expressing *Photinus pyralis* (firefly) *Renilla reniformis* (sea pansy) luciferases and with dsRNA 500mers (400ng), either homologous to firefly luciferase mRNA (dsLUC) or non-homologous (dsGFP). Dual luciferase assays were carried out using an Analytical Scientific Instruments model 3010 Luminometer. In this assay *Renilla* luciferase serves as an internal control for dsRNA-specific suppression of firefly luciferase activity. These data demonstrate that 500mer dsRNA can specifically suppress cognate gene expression in vivo.

35 Figure 29: P19 (a pluripotent, mouse embryonic cell line) cells transfected with plasmids expressing *Photinus pyralis* (firefly) *Renilla reniformis* (sea pansy) luciferases and with dsRNA 500mers (500ng), either homologous to firefly luciferase mRNA (dsLUC) or non-homologous (dsGFP). Dual luciferase assays were carried out using an Analytical Scientific Instruments model 3010 Luminometer. In this assay *Renilla*

luciferase serves as an internal control for dsRNA-specific suppression of firefly luciferase activity. These data further demonstrate that 500mer dsRNA can specifically suppress cognate gene expression in vivo and that the effect is stable over time .

5 Figure 30: S10 fractions from P19 cell lysates were used for in vitro translations of mRNA coding for Photinus pyralis (firefly) Renilla reniformis (sea pansy) luciferases. Translation reactions were programmed with various amounts of dsRNA 500mers, either homologous to firefly luciferase mRNA (dsLUC) or non-homologous (dsGFP). Reactions were carried out at 30 degrees for 1 hour, after which dual luciferase assays were carried out using an Analytical Scientific Instruments model 3010 Luminometer. In  
10 this assay Renilla luciferase serves as an internal control for dsRNA-specific suppression of firefly luciferase activity. These data demonstrate that 500mer dsRNA can specifically suppress cognate gene expression in vitro in a manner consistent with post-transcriptional gene silencing. Anti-sense firefly RNA did not differ significantly from dsGFP control (approximately 10%) (data not shown).

15 Figure 31: S10 fractions from P19 cell lysates were used for in vitro translations of mRNA coding for Photinus pyralis (firefly) Renilla reniformis (sea pansy) luciferases. Translation reactions were programmed with dsRNA or asRNA 500mers, either complementary to firefly luciferase mRNA (asLUC and dsLUC) or non-complementary (dsGFP). Reactions were carried out at 30 degrees for 1 hour, after a 30min preincubation  
20 with dsRNA or asRNA. Dual luciferase assays were carried out using an Analytical Scientific Instruments model 3010 Luminometer. In this assay Renilla luciferase serves as an internal control for dsRNA-specific suppression of firefly luciferase activity. These data demonstrate that 500mer double-stranded RNA (dsRNA) but not anti-sense RNA (asRNA) suppresses cognate gene expression in vitro in a manner consistent with post-  
25 transcriptional gene silencing.

Figure 32: P19 cells were grown in 6-well tissue culture plates to approximately 60% confluence. Various amounts of dsRNA, either homologous to firefly luciferase mRNA (dsLUC) or non-homologous (dsGFP), were added to each well and incubated for 12hrs under normal tissue culture conditions. Cells were then transfected with plasmids  
30 expressing Photinus pyralis (firefly) Renilla reniformis (sea pansy) luciferases and with dsRNA 500mers (500ng). Dual luciferase assays were carried out 12hrs post-transfection using an Analytical Scientific Instruments model 3010 Luminometer. In this assay Renilla luciferase serves as an internal control for dsRNA-specific suppression of firefly luciferase activity. These data show that 500mer dsRNA can specifically suppress cognate gene  
35 expression in vivo without transfection under normal tissue culture conditions.

Figure 33: Is a graph illustrating the relative rate of expression luciferase in cells which are treated with various antisense and dsRNA constructs.

## **Detailed Description of the Certain Preferred Embodiments**

### *I. Overview*

5 The present invention provides methods for attenuating gene expression in a cell using gene-targeted double stranded RNA (dsRNA). The dsRNA contains a nucleotide sequence that hybridizes under physiologic conditions of the cell to the nucleotide sequence of at least a portion of the gene to be inhibited (the “target” gene).

A significant aspect to certain embodiments of the present invention relates to the demonstration in the present application that RNAi can in fact be accomplished in cultured 10 cells, rather than whole organisms as described in the art.

Another salient feature of the present invention concerns the ability to carry out RNAi in higher eukaryotes, particularly in non-oocytic cells of mammals, e.g., cells from adult mammals as an example.

As described in further detail below, the present invention(s) are based on the 15 discovery that the RNAi phenomenon is mediated by a set of enzyme activities, including an essential RNA component, that are evolutionarily conserved in eukaryotes ranging from plants to mammals.

One enzyme contains an essential RNA component. After partial purification, a multi-component nuclease (herein “RISC nuclease”) co-fractionates with a discrete, 22- 20 nucleotide RNA species which may confer specificity to the nuclease through homology to the substrate mRNAs. The short RNA molecules are generated by a processing reaction from the longer input dsRNA. Without wishing to be bound by any particular theory, these 22mer guide RNAs may serve as guide sequences that instruct the RISC nuclease to destroy specific mRNAs corresponding to the dsRNA sequences.

As illustrated in Figure 33, double stranded forms of the 22-mer guide RNA can be 25 sufficient in length to induce sequence-dependent dsRNA inhibition of gene expression. In the illustrated example, dsRNA constructs are administered to cells having a recombinant luciferase reporter gene. The control cell, e.g., no exogenously added RNA, the level of expression of the luciferase reporter is normalized to be the value of “1”. As 30 illustrated, both long (500-mer) and short (22-mer) dsRNA constructs complementary to the luciferase gene could inhibit expression of that gene product relative to the control cell. On the other hand, similarly sized dsRNA complementary to the coding sequence for another protein, green fluorescence protein (GFP), did not significantly effect the expression of luciferase –indicating that the inhibitory phenomena was in each case 35 sequence-dependent. Likewise, single stranded 22-mers of luciferase did not inhibit

expression of that gene – indicating that the inhibitory phenomena is double stranded-dependent.

The appended examples also identify an enzyme, Dicer, that can produce the putative guide RNAs. Dicer is a member of the RNase III family of nucleases that specifically cleave dsRNA and is evolutionarily conserved in worms, flies, plants, fungi and, as described herein, mammals. The enzyme has a distinctive structure which includes a helicase domain and dual RNase III motifs. Dicer also contains a region of homology to the RDE1/QDE2/ARGONAUTE family, which have been genetically linked to RNAi in lower eukaryotes. Indeed, activation of, or overexpression of Dicer may be sufficient in many cases to permit RNA interference in otherwise non-receptive cells, such as cultured eukaryotic cells, or mammalian (non-oocytic) cells in culture or in whole organisms.

In certain embodiments, the cells can be treated with an agent(s) that inhibits the general double-stranded RNA response(s) by the host cells, such as may give rise to sequence-independent apoptosis. For instance, the cells can be treated with agents that inhibit the dsRNA-dependent protein kinase known as PKR (protein kinase RNA-activated). Double stranded RNAs in mammalian cells typically activate protein kinase PKR and leads to apoptosis. The mechanism of action of PKR includes phosphorylation and inactivation eIF2a (Fire (1999) Trends Genet 15:358). It has also been reported that induction of NF- $\kappa$ B by PKR is involved in apoptosis commitment and this process is mediated through activation of the IKK complex. This sequence-independent response may reflect a form of primitive immune response, since the presence of dsRNA is a common feature of many viral lifecycles.

As described herein, Applicants have demonstrated that the PKR response can be overcome in favor of the sequence-specific RNAi response. However, in certain instances, it can be desirable to treat the cells with agents which inhibit expression of PKR, cause its destruction, and/or inhibit the kinase activity of PKF are specifically contemplated for use in the present method. Likewise, overexpression of or agents which ectopically activate IF2 $\alpha$  can be used. Other agents which can be used to suppress the PKR response include inhibitors of IKK phosphorylation of I $\kappa$ B, inhibitors of I $\kappa$ B ubiquitination, inhibitors of I $\kappa$ B degradation, inhibitors of NF- $\kappa$ B nuclear translocation, and inhibitors of NF- $\kappa$ B interaction with  $\kappa$ B response elements.

Other inhibitors of sequence-independent dsRNA response in cells include the gene product of the vaccinia virus E3L. The E3L gene product contains two distinct domains. A conserved carboxy-terminal domain has been shown to bind double-stranded RNA (dsRNA) and inhibit the antiviral dsRNA response by cells. Expression of at least that portion of the E3L gene in the host cell, or the use of polypeptide or peptidomimetics thereof, can be used to suppress the general dsRNA response. Caspase inhibitors

sensitized cells to killing by double-stranded RNA. Accordingly, ectopic expression or activated of caspases in the host cell can be used to suppress the general dsRNA response.

In other embodiments, the subject method is carried out in cells which have little or no general response to double stranded RNA, e.g., have no PKR-dependent dsRNA response, at least under the culture conditions. As illustrated in Figures 28-32, CHO and P19 cells can be used without having to inhibit PKR or other general dsRNA responses.

Thus, the present invention provides a process and compositions for inhibiting expression of a target gene in a cell, especially a mammalian cell. In certain embodiments, the process comprises introduction of RNA (the "dsRNA construct") with partial or fully double-stranded character into the cell or into the extracellular environment. Inhibition is specific in that a nucleotide sequence from a portion of the target gene is chosen to produce the dsRNA construct. In preferred embodiments, the method utilizes a cell in which Dicer and/or Argonaute activities are recombinantly expressed or otherwise ectopically activated. This process can be (1) effective in attenuating gene expression, (2) specific to the targeted gene, and (3) general in allowing inhibition of many different types of target gene.

## *II. Definitions*

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to that it has been linked. One type of vector is a genomic integrated vector, or "integrated vector", which can become integrated into the chromosomal DNA of the host cell. Another type of vector is an episomal vector, i.e., a nucleic acid capable of extra-chromosomal replication. Vectors capable of directing the expression of genes to that they are operatively linked are referred to herein as "expression vectors". In the present specification, "plasmid" and "vector" are used interchangeably unless otherwise clear from the context.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a polypeptide of the present invention, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding such regulatory polypeptides, that may optionally include intron

sequences that are derived from chromosomal DNA. The term "intron" refers to a DNA  
 sequence present in a given gene that is not translated into protein and is generally found  
 between exons. As used herein, the term "transfection" means the introduction of a  
 nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene  
 5 transfer.

A "protein coding sequence" or a sequence that "encodes" a particular polypeptide  
 or peptide, is a nucleic acid sequence that is transcribed (in the case of DNA) and is  
 translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under  
 the control of appropriate regulatory sequences. The boundaries of the coding sequence  
 10 are determined by a start codon at the 5' (amino) terminus and a translation stop codon at  
 the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA  
 from procaryotic or eukaryotic mRNA, genomic DNA sequences from procaryotic or  
 eukaryotic DNA, and even synthetic DNA sequences. A transcription termination  
 sequence will usually be located 3' to the coding sequence.

15 Likewise, "encodes", unless evident from its context, will be meant to include  
 DNA sequences that encode a polypeptide, as the term is typically used, as well as DNA  
 sequences that are transcribed into inhibitory antisense molecules.

The term "loss-of-function", as it refers to genes inhibited by the subject RNAi  
 method, refers a diminishment in the level of expression of a gene when compared to the  
 20 level in the absense of dsRNA constructs.

The term "expression" with respect to a gene sequence refers to transcription of the  
 gene and, as appropriate, translation of the resulting mRNA transcript to a protein. Thus,  
 as will be clear from the context, expression of a protein coding sequence results from  
 transcription and translation of the coding sequence.

25 "Cells," "host cells" or "recombinant host cells" are terms used interchangeably  
 herein. It is understood that such terms refer not only to the particular subject cell but to  
 the progeny or potential progeny of such a cell. Because certain modifications may occur  
 in succeeding generations due to either mutation or environmental influences, such  
 progeny may not, in fact, be identical to the parent cell, but are still included within the  
 30 scope of the term as used herein.

The term "cultured cells" refers to cells suspended in culture, e.g., dispersed in  
 culture or in the form tissue. It does not, however, include oocytes or whole embryos  
 (including blastocysts and the like) which may be provided in culture. In certain  
 embodiments, the cultured cells are adults cells, e.g., non-embryonic.

35 By "recombinant virus" is meant a virus that has been genetically altered, e.g., by  
 the addition or insertion of a heterologous nucleic acid construct into the particle.

FOR PUBLICATION

As used herein, the terms "transduction" and "transfection" are art recognized and mean the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a dsRNA construct.

"Transient transfection" refers to cases where exogenous DNA does not integrate into the genome of a transfected cell, e.g., where episomal DNA is transcribed into mRNA and translated into protein.

A cell has been "stably transfected" with a nucleic acid construct when the nucleic acid construct is capable of being inherited by daughter cells.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter gene" operatively linked to at least one transcriptional regulatory sequence. Transcription of the reporter gene is controlled by these sequences to which they are linked. The activity of at least one or more of these control sequences can be directly or indirectly regulated by the target receptor protein. Exemplary transcriptional control sequences are promoter sequences. A reporter gene is meant to include a promoter-reporter gene construct that is heterologously expressed in a cell.

As used herein, "transformed cells" refers to cells that have spontaneously converted to a state of unrestrained growth, i.e., they have acquired the ability to grow through an indefinite number of divisions in culture. Transformed cells may be characterized by such terms as neoplastic, anaplastic and/or hyperplastic, with respect to their loss of growth control. For purposes of this invention, the terms "transformed phenotype of malignant mammalian cells" and "transformed phenotype " are intended to encompass, but not be limited to, any of the following phenotypic traits associated with cellular transformation of mammalian cells: immortalization, morphological or growth transformation, and tumorigenicity, as detected by prolonged growth in cell culture, growth in semi-solid media, or tumorigenic growth in immuno-incompetent or syngeneic animals.

As used herein, "proliferating" and "proliferation" refer to cells undergoing mitosis.

As used herein, "immortalized cells" refers to cells that have been altered via chemical, genetic, and/or recombinant means such that the cells have the ability to grow through an indefinite number of divisions in culture.

The "growth state" of a cell refers to the rate of proliferation of the cell and the state of differentiation of the cell.



*III. Exemplary embodiments of Isolation Method*

One aspect of the invention provides a method for potentiating RNAi by induction or ectopic activation of an RNAi enzyme in a cell (in vivo or in vitro) or cell-free mixtures. In preferred embodiments, the RNAi activity is activated or added to a mammalian cell, e.g., a human cell, which cell may be provided in vitro or as part of a whole organism. In other embodiments, the subject method is carried out using eukaryotic cells generally (except for oocytes) in culture. For instance, the Dicer enzyme may be activated by virtue of being recombinantly expressed or it may be activated by use of an agent which (i) induces expression of the endogenous gene, (ii) stabilizes the protein from degradation, and/or (iii) allosterically modifies the enzyme to increase its activity (by altering its Kcat, Km or both).

A. Dicer and Argonaut Activities

In certain embodiment, at least one of the activated RNAi enzymes is Dicer, or a homolog thereof. In certain preferred embodiments, the present method provides for ectopic activation of Dicer. As used herein, the term “Dicer” refers to a protein which (a) mediates an RNAi response and (b) has an amino acid sequence at least 50 percent identical, and more preferably at least 75, 85, 90 or 95 percent identical to SEQ ID No. 2 or 4, and/or which can be encoded by a nucleic acid which hybridizes under wash conditions of 2 x SSC at 22°C, and more preferably 0.2 x SSC at 65°C, to a nucleotide represented by SEQ ID No. 1 or 3. Accordingly, the method may comprise introducing a dsRNA construct into a cell in which Dicer has been recombinantly expressed or otherwise ectopically activated.

In certain embodiment, at least one of the activated RNAi enzymes is Argonaut, or a homolog thereof. In certain preferred embodiments, the present method provides for ectopic activation of Argonaut. As used herein, the term “Argonaut” refers to a protein which (a) mediates an RNAi response and (b) has an amino acid sequence at least 50 percent identical, and more preferably at least 75, 85, 90 or 95 percent identical to the amino acid sequence shown in Figure 24. Accordingly, the method may comprise introducing a dsRNA construct into a cell in which Argonaut has been recombinantly expressed or otherwise ectopically activated.

This invention also provides expression vectors containing a nucleic acid encoding a Dicer or Argonaut polypeptides, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the

subject Dicer or Argonaut proteins. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences, sequences that control the expression of a DNA sequence when operatively linked to it, may be used in these vectors to express DNA sequences encoding Dicer or Argonaut polypeptides of this invention. Such useful expression control sequences, include, for example, a (100) LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage  $\lambda$ , the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast  $\alpha$ -mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed.

Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered.

The recombinant Dicer or Argonaut genes can be produced by ligating nucleic acid encoding a Dicer or Argonaut polypeptide into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject Dicer or Argonaut polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a Dicer or Argonaut polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as E. coli.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach et al. (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in *E. coli* due the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance

markers such as ampicillin can be used. In an illustrative embodiment, a Dicer or Argonaut polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of a Dicer or Argonaut gene.

The preferred mammalian expression vectors contain both prokaryotic sequences, 5 to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNA1/amp, pcDNA1/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from 10 bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well 15 known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

In yet another embodiment, the subject invention provides a “gene activation” 20 construct which, by homologous recombination with a genomic DNA, alters the transcriptional regulatory sequences of an endogenous Dicer or Argonaut gene. For instance, the gene activation construct can replace the endogenous promoter of a Dicer or Argonaut gene with a heterologous promoter, e.g., one which causes constitutive expression of the Dicer or Argonaut gene or which causes inducible expression of the gene 25 under conditions different from the normal expression pattern of Dicer or Argonaut. A variety of different formats for the gene activation constructs are available. See, for example, the Transkaryotic Therapies, Inc PCT publications WO93/09222, WO95/31560, WO96/29411, WO95/31560 and WO94/12650.

In preferred embodiments, the nucleotide sequence used as the gene activation 30 construct can be comprised of (1) DNA from some portion of the endogenous Dicer or Argonaut gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) heterologous transcriptional regulatory sequence(s) which is to be operably linked to the coding sequence for the genomic Dicer or Argonaut gene upon recombination of the gene activation construct. For use in generating cultures of Dicer or 35 Argonaut producing cells, the construct may further include a reporter gene to detect the presence of the knockout construct in the cell.

The gene activation construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to provide the heterologous regulatory sequences in operative association with the native Dicer or Argonaut gene. Such insertion occurs by homologous recombination, i.e., recombination regions of the activation
 5 construct that are homologous to the endogenous Dicer or Argonaut gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA.

The terms "recombination region" or "targeting sequence" refer to a segment (i.e., a portion) of a gene activation construct having a sequence that is substantially identical to
 10 or substantially complementary to a genomic gene sequence, e.g., including 5' flanking sequences of the genomic gene, and can facilitate homologous recombination between the genomic sequence and the targeting transgene construct.

As used herein, the term "replacement region" refers to a portion of a activation construct which becomes integrated into an endogenous chromosomal location following
 15 homologous recombination between a recombination region and a genomic sequence.

The heterologous regulatory sequences, e.g., which are provided in the replacement region, can include one or more of a variety elements, including: promoters (such as constitutive or inducible promoters), enhancers, negative regulatory elements, locus control regions, transcription factor binding sites, or combinations thereof.

Promoters/enhancers which may be used to control the expression of the targeted gene *in vivo* include, but are not limited to, the cytomegalovirus (CMV) promoter/enhancer (Karasuyama et al., 1989, *J. Exp. Med.*, 169:13), the human  $\beta$ -actin promoter (Gunning et al. (1987) *PNAS* 84:4831-4835), the glucocorticoid-inducible promoter present in the mouse mammary tumor virus long terminal repeat (MMTV LTR)
 20 (Klessig et al. (1984) *Mol. Cell Biol.* 4:1354-1362), the long terminal repeat sequences of Moloney murine leukemia virus (MuLV LTR) (Weiss et al. (1985) *RNA Tumor Viruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), the SV40 early or late region promoter (Bernoist et al. (1981) *Nature* 290:304-310; Templeton et al. (1984) *Mol. Cell Biol.*, 4:817; and Sprague et al. (1983) *J. Virol.*, 45:773), the promoter contained in
 25 the 3' long terminal repeat of Rous sarcoma virus (RSV) (Yamamoto et al., 1980, *Cell*, 22:787-797), the herpes simplex virus (HSV) thymidine kinase promoter/enhancer (Wagner et al. (1981) *PNAS* 82:3567-71), and the herpes simplex virus LAT promoter (Wolfe et al. (1992) *Nature Genetics*, 1:379-384).

In still other embodiments, the replacement region merely deletes a negative transcriptional control element of the native gene, e.g., to activate expression, or ablates a
 35 positive control element, e.g., to inhibit expression of the targeted gene.

## B. Cell/Organism

The cell with the target gene may be derived from or contained in any organism (e.g., plant, animal, protozoan, virus, bacterium, or fungus). The dsRNA construct may be synthesized either in vivo or in vitro. Endogenous RNA polymerase of the cell may mediate transcription in vivo, or cloned RNA polymerase can be used for transcription in vivo or in vitro. For generating double stranded transcripts from a transgene in vivo, a regulatory region may be used to transcribe the RNA strand (or strands).

Furthermore, genetic manipulation becomes possible in organisms that are not classical genetic models. Breeding and screening programs may be accelerated by the ability to rapidly assay the consequences of a specific, targeted gene disruption. Gene disruptions may be used to discover the function of the target gene, to produce disease models in which the target gene are involved in causing or preventing a pathological condition, and to produce organisms with improved economic properties.

The cell with the target gene may be derived from or contained in any organism. The organism may be a plant, animal, protozoan, bacterium, virus, or fungus. The plant may be a monocot, dicot or gymnosperm; the animal may be a vertebrate or invertebrate. Preferred microbes are those used in agriculture or by industry, and those that are pathogenic for plants or animals. Fungi include organisms in both the mold and yeast morphologies.

Plants include arabidopsis; field crops (e.g., alfalfa, barley, bean, corn, cotton, flax, pea, rape, rice, rye, safflower, sorghum, soybean, sunflower, tobacco, and wheat); vegetable crops (e.g., asparagus, beet, broccoli, cabbage, carrot, cauliflower, celery, cucumber, eggplant, lettuce, onion, pepper, potato, pumpkin, radish, spinach, squash, taro, tomato, and zucchini); fruit and nut crops (e.g., almond, apple, apricot, banana, blackberry, blueberry, cacao, cherry, coconut, cranberry, date, fava, filbert, grape, grapefruit, guava, kiwi, lemon, lime, mango, melon, nectarine, orange, papaya, passion fruit, peach, peanut, pear, pineapple, pistachio, plum, raspberry, strawberry, tangerine, walnut, and watermelon); and ornamentals (e.g., alder, ash, aspen, azalea, birch, boxwood, camellia, carnation, chrysanthemum, elm, fir, ivy, jasmine, juniper, oak, palm, poplar, pine, redwood, rhododendron, rose, and rubber).

Examples of vertebrate animals include fish, mammal, cattle, goat, pig, sheep, rodent, hamster, mouse, rat, primate, and human.

Invertebrate animals include nematodes, other worms, drosophila, and other insects. Representative genera of nematodes include those that infect animals (e.g., Ancylostoma, Ascaridia, Ascaris, Bunostomum, Caenorhabditis, Capillaria, Chabertia, Cooperia, Dictyocaulus, Haemonchus, Heterakis, Nematodirus, Oesophagostomum,

Ostertagia, Oxyuris, Parascaris, Strongylus, Toxascaris, Trichuris, Trichostrongylus, Tfllichonema, Toxocara, Uncinaria) and those that infect plants (e.g., B ursaphalenchus, Criconerriella, Diiylenchus, Ditylenchus, Globodera, Helicotylenchus, Heterodera, Longidorus, Melodoigyne, Nacobbus, Paratylenchus, Pratylenchus, Radopholus, 5 Rotelynychus, Tylenchus, and Xiphinerna). Representative orders of insects include Coleoptera, Diptera, Lepidoptera, and Homoptera.

The cell having the target gene may be from the germ line or somatic, totipotent or pluripotent, dividing or non-dividing, parenchyma or epithelium, immortalized or transformed, or the like. The cell may be a stem cell or a differentiated cell. Cell types that 10 are differentiated include adipocytes, fibroblasts, myocytes, cardiomyocytes, endothelium, neurons, glia, blood cells, megakaryocytes, lymphocytes, macrophages, neutrophils, eosinophils, basophils, mast cells, leukocytes, granulocytes, keratinocytes, chondrocytes, osteoblasts, osteoclasts, hepatocytes, and cells of the endocrine or exocrine glands.

### 15 C. Targeted Genes

The target gene may be a gene derived from the cell, an endogenous gene, a transgene, or a gene of a pathogen which is present in the cell after infection thereof. Depending on the particular target gene and the dose of double stranded RNA material delivered, the procedure may provide partial or complete loss of function for the target 20 gene. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells. Quantitation of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein.

“Inhibition of gene expression” refers to the absence (or observable decrease) in 25 the level of protein and/or mRNA product from a target gene. “Specificity” refers to the ability to inhibit the target gene without manifest effects on other genes of the cell. The consequences of inhibition can be confirmed by examination of the outward properties of the cell or organism (as presented below in the examples) or by biochemical techniques such as RNA solution hybridization, nuclease protection, Northern hybridization, reverse 30 transcription, gene expression monitoring with a microarray, antibody binding, enzyme linked immunosorbent assay (ELISA), Western blotting, radiolimmunoassay (RIA), other immunoassays, and fluorescence activated cell analysis (FACS). For RNA-mediated inhibition in a cell line or whole organism, gene expression is conveniently assayed by use of a reporter or drug resistance gene whose protein product is easily assayed. Such reporter 35 genes include acetohydroxyacid synthase (AHAS), alkaline phosphatase (AP), beta galactosidase (LacZ), beta glucuronidase (GUS), chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), horseradish peroxidase (HRP), luciferase (Luc),

nopaline synthase (NOS), octopine synthase (OCS), and derivatives thereof multiple selectable markers are available that confer resistance to ampicillin, bleomycin, chloramphenicol, gentamycin, hygromycin, kanamycin, lincomycin, methotrexate, phosphinothricin, puromycin, and tetracyclin.

5            Depending on the assay, quantitation of the amount of gene expression allows one to determine a degree of inhibition which is greater than 10%, 33%, 50%, 90%, 95% or 99% as compared to a cell not treated according to the present invention. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells (e.g., at least 10%, 20%, 50%, 75%, 90%, or 95% of targeted  
10 cells). Quantitation of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein. As an example, the efficiency of inhibition may be determined by assessing the amount of gene product in the cell: mRNA may be detected with a hybridization probe having a nucleotide sequence outside the region used for the inhibitory double-stranded RNA, or translated polypeptide  
15 may be detected with an antibody raised against the polypeptide sequence of that region.

As disclosed herein, the present invention may is not limited to any type of target gene or nucleotide sequence. But the following classes of possible target genes are listed for illustrative purposes: developmental genes (e.g., adhesion molecules, cyclin kinase inhibitors, Writ family members, Pax family members, Winged helix family members,  
20 Hox family members, cytokines/lymphokines and their receptors, growth/differentiation factors and their receptors, neurotransmitters and their receptors); oncogenes (e.g., ABLI, BCLI, BCL2, BCL6, CBFA2, CBL, CSFIR, ERBA, ERBB, EBRB2, ETSI, ETS1, ETV6, FGR, FOS, FYN, HCR, HRAS, JUN, KRAS, LCK, LYN, MDM2, MLL, MYB, MYC, MYCLI, MYCN, NRAS, PIM 1, PML, RET, SRC, TALI, TCL3, and YES); tumor  
25 suppressor genes (e.g., APC, BRCA 1, BRCA2, MADH4, MCC, NF 1, NF2, RB 1, TP53, and WTI); and enzymes (e.g., ACC synthases and oxidases, ACP desaturases and hydroxylases, ADP-glucose pyrophorylases, ATPases, alcohol dehydrogenases, amylases, amyloglucosidases, catalases, cellulases, chalcone synthases, chitinases, cyclooxygenases, decarboxylases, dextrinases, DNA and RNA polymerases, galactosidases, glucanases,  
30 glucose oxidases, granule-bound starch synthases, GTPases, helicases, hemicellulases, integrases, inulinases, invertases, isomerases, kinases, lactases, lipases, lipoxygenases, lysozymes, nopaline synthases, octopine synthases, pectinesterases, peroxidases, phosphatases, phospholipases, phosphorylases, phytases, plant growth regulator synthases, polygalacturonases, proteinases and peptidases, pullanases, recombinases, reverse  
35 transcriptases, RUBISCOs, topoisomerases, and xylanases).

#### D. dsRNA constructs

The dsRNA construct may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the nucleoside. For example, the phosphodiester linkages of natural RNA may be modified to include at least one of a nitrogen or sulfur heteroatom. Modifications in RNA structure may be tailored to allow specific genetic inhibition while avoiding a general panic response in some organisms which is generated by dsRNA. Likewise, bases may be modified to block the activity of adenosine deaminase. The dsRNA construct may be produced enzymatically or by partial/total organic synthesis, any modified ribonucleotide can be introduced by in vitro enzymatic or organic synthesis.

The dsRNA construct may be directly introduced into the cell (i.e., intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, introduced orally, or may be introduced by bathing an organism in a solution containing RNA. Methods for oral introduction include direct mixing of RNA with food of the organism, as well as engineered approaches in which a species that is used as food is engineered to express an RNA, then fed to the organism to be affected. Physical methods of introducing nucleic acids include injection directly into the cell or extracellular injection into the organism of an RNA solution.

The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses (e.g., at least 5, 10, 100, 500 or 1000 copies per cell) of double-stranded material may yield more effective inhibition; lower doses may also be useful for specific applications. Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition.

dsRNA constructs containing a nucleotide sequences identical to a portion of the target gene are preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may be optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, Sequence Analysis Primer, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene is preferred. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion



of the target gene transcript (e.g., 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50°C or 70°C hybridization for 12-16 hours; followed by washing). In certain preferred embodiments, the length of the dsRNA is at least 20, 21 or 22 nucleotides in length, e.g., corresponding in size to RNA products produced by Dicer-dependent cleavage. In certain  
5 embodiments, the dsRNA construct is at least 25, 50, 100, 200, 300 or 400 bases. In certain embodiments, the dsRNA construct is 400-800 bases in length.

100% sequence identity between the RNA and the target gene is not required to practice the present invention. Thus the invention has the advantage of being able to tolerate sequence variations that might be expected due to genetic mutation, strain  
10 polymorphism, or evolutionary divergence.

The dsRNA construct may be synthesized either in vivo or in vitro. Endogenous RNA polymerase of the cell may mediate transcription in vivo, or cloned RNA polymerase can be used for transcription in vivo or in vitro. For transcription from a transgene in vivo or an expression construct, a regulatory region (e.g., promoter, enhancer, silencer, splice donor and acceptor, polyadenylation) may be used to transcribe the dsRNA  
15 strand (or strands). Inhibition may be targeted by specific transcription in an organ, tissue, or cell type; stimulation of an environmental condition (e.g., infection, stress, temperature, chemical inducers); and/or engineering transcription at a developmental stage or age. The RNA strands may or may not be polyadenylated; the RNA strands may or may not be capable of being translated into a polypeptide by a cell's translational apparatus. The dsRNA construct may be chemically or enzymatically synthesized by manual or automated reactions. The dsRNA construct may be synthesized by a cellular RNA  
20 polymerase or a bacteriophage RNA polymerase (e.g., T3, T7, SP6). The use and production of an expression construct are known in the art 32,33,34 (see also WO 97/32016; U.S. Pat. Nos. 5,593,874, 5,698,425, 5,712,135, 5,789,214, and 5,804,693; and the references cited therein). If synthesized chemically or by in vitro enzymatic synthesis, the RNA may be purified prior to introduction into the cell. For example, RNA can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography or a combination thereof. Alternatively, the dsRNA  
30 construct may be used with no or a minimum of purification to avoid losses due to sample processing.. The dsRNA construct may be dried for storage or dissolved in an aqueous solution. The solution may contain buffers or salts to promote annealing, and/or stabilization of the duplex strands.

Physical methods of introducing nucleic acids include injection of a solution  
35 containing the dsRNA construct, bombardment by particles covered by the dsRNA construct, soaking the cell or organism in a solution of the RNA, or electroporation of cell membranes in the presence of the dsRNA construct. A viral construct packaged into a viral

particle would accomplish both efficient introduction of an expression construct into the cell and transcription of dsRNA construct encoded by the expression construct. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemical-mediated transport, such as calcium phosphate, and the like. Thus the dsRNA construct may be introduced along with components that perform one or more of the following activities: enhance RNA uptake by the cell, promote annealing of the duplex strands, stabilize the annealed strands, or other-wise increase inhibition of the target gene.

10           E. Illustrative Uses

One utility of the present invention is as a method of identifying gene function in an organism, especially higher eukaryotes comprising the use of double-stranded RNA to inhibit the activity of a target gene of previously unknown function. Instead of the time consuming and laborious isolation of mutants by traditional genetic screening, functional genomics would envision determining the function of uncharacterized genes by employing the invention to reduce the amount and/or alter the timing of target gene activity. The invention could be used in determining potential targets for pharmaceuticals, understanding normal and pathological events associated with development, determining signaling pathways responsible for postnatal development/aging, and the like. The increasing speed of acquiring nucleotide sequence information from genomic and expressed gene sources, including total sequences for mammalian genomes, can be coupled with the invention to determine gene function in a cell or in a whole organism. The preference of different organisms to use particular codons, searching sequence databases for related gene products, correlating the linkage map of genetic traits with the physical map from which the nucleotide sequences are derived, and artificial intelligence methods may be used to define putative open reading frames from the nucleotide sequences acquired in such sequencing projects.

A simple assay would be to inhibit gene expression according to the partial sequence available from an expressed sequence tag (EST). Functional alterations in growth, development, metabolism, disease resistance, or other biological processes would be indicative of the normal role of the EST's gene product.

The ease with which the dsRNA construct can be introduced into an intact cell/organism containing the target gene allows the present invention to be used in high throughput screening (HTS). For example, duplex RNA can be produced by an amplification reaction using primers flanking the inserts of any gene library derived from the target cell or organism. Inserts may be derived from genomic DNA or mRNA (e.g., cDNA and cRNA). Individual clones from the library can be replicated and then isolated

in separate reactions, but preferably the library is maintained in individual reaction vessels (e.g., a 96 well microtiter plate) to minimize the number of steps required to practice the invention and to allow automation of the process.

In an exemplary embodiment, the subject invention provides an arrayed library of RNAi constructs. The array may in the form of solutions, such as multi-well plates, or may be “printed” on solid substrates upon which cells can be grown. To illustrate, solutions containing duplex RNAs that are capable of inhibiting the different expressed genes can be placed into individual wells positioned on a microtiter plate as an ordered array, and intact cells/organisms in each well can be assayed for any changes or modifications in behavior or development due to inhibition of target gene activity.

In one embodiment, the subject method uses an arrayed library of RNAi constructs to screen for combinations of RNAi that is lethal to host cells. Synthetic lethality is a bedrock principle of experimental genetics. A synthetic lethality describes the properties of two mutations which, individually, are tolerated by the organism but which, in combination, are lethal. The subject arrays can be used to identify loss-of-function mutations that are lethal in combination with alterations in other genes, such as activated oncogenes or loss-of-function mutations to tumor suppressors. To achieve this, one can create “phenotype arrays” using cultured cells. Expression of each of a set of genes, such as the host cell’s genome, can be individually systematically disrupted using RNA interference. Combination with alterations in oncogene and tumor suppressor pathways can be used to identify synthetic lethal interactions that may identify novel therapeutic targets.

In certain embodiments, the RNAi constructs can be fed directly to, injected into, the cell/organism containing the target gene. Alternatively, the duplex RNA can be produced by in vivo or in vitro transcription from an expression construct used to produce the library. The construct can be replicated as individual clones of the library and transcribed to produce the RNA; each clone can then be fed to, or injected into, the cell/organism containing the target gene. The function of the target gene can be assayed from the effects it has on the cell/organism when gene activity is inhibited. This screening could be amenable to small subjects that can be processed in large number, for example, tissue culture cells derived from mammals, especially primates, and most preferably humans.

If a characteristic of an organism is determined to be genetically linked to a polymorphism through RFLP or QTL analysis, the present invention can be used to gain insight regarding whether that genetic polymorphism might be directly responsible for the characteristic. For example, a fragment defining the genetic polymorphism or sequences in the vicinity of such a genetic polymorphism can be amplified to produce an RNA, the

duplex RNA can be introduced to the organism or cell, and whether an alteration in the characteristic is correlated with inhibition can be determined. Of course, there may be trivial explanations for negative results with this type of assay, for example: inhibition of the target gene causes lethality, inhibition of the target gene may not result in any observable alteration, the fragment contains nucleotide sequences that are not capable of inhibiting the target gene, or the target gene's activity is redundant.

The present invention may be useful in allowing the inhibition of essential genes. Such genes may be required for cell or organism viability at only particular stages of development or cellular compartments. The functional equivalent of conditional mutations may be produced by inhibiting activity of the target gene when or where it is not required for viability. The invention allows addition of RNA at specific times of development and locations in the organism without introducing permanent mutations into the target genome.

If alternative splicing produced a family of transcripts that were distinguished by usage of characteristic exons, the present invention can target inhibition through the appropriate exons to specifically inhibit or to distinguish among the functions of family members. For example, a hormone that contained an alternatively spliced transmembrane domain may be expressed in both membrane bound and secreted forms. Instead of isolating a nonsense mutation that terminates translation before the transmembrane domain, the functional consequences of having only secreted hormone can be determined according to the invention by targeting the exon containing the transmembrane domain and thereby inhibiting expression of membrane-bound hormone.

The present invention may be used alone or as a component of a kit having at least one of the reagents necessary to carry out the in vitro or in vivo introduction of RNA to test samples or subjects. Preferred components are the dsRNA and a vehicle that promotes introduction of the dsRNA. Such a kit may also include instructions to allow a user of the kit to practice the invention.

Alternatively, an organism may be engineered to produce dsRNA which produces commercially or medically beneficial results, for example, resistance to a pathogen or its pathogenic effects, improved growth, or novel developmental patterns.

*IV. Exemplification*

The invention, now being generally described, will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit the invention.

### Example 1: An RNA-directed nuclease mediates RNAi gene silencing

In a diverse group of organisms that includes *Caenorhabditis elegans*, *Drosophila*, planaria, hydra, trypanosomes, fungi and plants, the introduction of double-stranded RNAs inhibits gene expression in a sequence-specific manner<sup>1-7</sup>. These responses, called RNA interference or post-transcriptional gene silencing, may provide anti-viral defence, modulate transposition or regulate gene expression<sup>1, 6, 8-10</sup>. We have taken a biochemical approach towards elucidating the mechanisms underlying this genetic phenomenon. Here we show that 'loss-of-function' phenotypes can be created in cultured *Drosophila* cells by transfection with specific double-stranded RNAs. This coincides with a marked reduction in the level of cognate cellular messenger RNAs. Extracts of transfected cells contain a nuclease activity that specifically degrades exogenous transcripts homologous to transfected double-stranded RNA. This enzyme contains an essential RNA component. After partial purification, the sequence-specific nuclease co-fractionates with a discrete, ~25-nucleotide RNA species which may confer specificity to the enzyme through homology to the substrate mRNAs.

Although double-stranded RNAs (dsRNAs) can provoke gene silencing in numerous biological contexts including *Drosophila*<sup>11, 12</sup>, the mechanisms underlying this phenomenon have remained mostly unknown. We therefore wanted to establish a biochemically tractable model in which such mechanisms could be investigated.

Transient transfection of cultured, *Drosophila* S2 cells with a *lacZ* expression vector resulted in  $\beta$ -galactosidase activity that was easily detectable by an *in situ* assay (Fig. 1a). This activity was greatly reduced by co-transfection with a dsRNA corresponding to the first 300 nucleotides of the *lacZ* sequence, whereas co-transfection with a control dsRNA (*CD8*) (Fig. 1a) or with single-stranded RNAs of either sense or antisense orientation (data not shown) had little or no effect. This indicated that dsRNAs could interfere, in a sequence-specific fashion, with gene expression in cultured cells.

To determine whether RNA interference (RNAi) could be used to target endogenous genes, we transfected S2 cells with a dsRNA corresponding to the first 540 nucleotides of *Drosophila cyclin E*, a gene that is essential for progression into S phase of the cell cycle. During log-phase growth, untreated S2 cells reside primarily in G2/M (Fig. 1b). Transfection with *lacZ* dsRNA had no effect on cell-cycle distribution, but transfection with the *cyclin E* dsRNA caused a G1-phase cell-cycle arrest (Fig. 1b). The ability of *cyclin E* dsRNA to provoke this response was length-dependent. Double-stranded RNAs of 540 and 400 nucleotides were quite effective, whereas dsRNAs of 200 and 300 nucleotides were less potent. Double-stranded *cyclin E* RNAs of 50 or 100 nucleotides were inert in our assay, and transfection with a single-stranded, antisense *cyclin E* RNA had virtually no effect.

One hallmark of RNAi is a reduction in the level of mRNAs that are homologous to the dsRNA. Cells transfected with the *cyclin E* dsRNA (bulk population) showed diminished endogenous *cyclin E* mRNA as compared with control cells (Fig. 1c). Similarly, transfection of cells with dsRNAs homologous to *fizzy*, a component of the anaphase-promoting complex (APC) or *cyclin A*, a cyclin that acts in S, G2 and M, also caused reduction of their cognate mRNAs (Fig. 1c). The modest reduction in *fizzy* mRNA levels in cells transfected with *cyclin A* dsRNA probably resulted from arrest at a point in the division cycle at which *fizzy* transcription is low<sup>14, 15</sup>. These results indicate that RNAi may be a generally applicable method for probing gene function in cultured *Drosophila* cells.

The decrease in mRNA levels observed upon transfection of specific dsRNAs into *Drosophila* cells could be explained by effects at transcriptional or post-transcriptional levels. Data from other systems have indicated that some elements of the dsRNA response may affect mRNA directly (reviewed in refs 1 and 6). We therefore sought to develop a cell-free assay that reflected, at least in part, RNAi.

S2 cells were transfected with dsRNAs corresponding to either *cyclin E* or *lacZ*. Cellular extracts were incubated with synthetic mRNAs of *lacZ* or *cyclin E*. Extracts prepared from cells transfected with the 540-nucleotide *cyclin E* dsRNA efficiently degraded the *cyclin E* transcript; however, the *lacZ* transcript was stable in these lysates (Fig. 2a). Conversely, lysates from cells transfected with the *lacZ* dsRNA degraded the *lacZ* transcript but left the *cyclin E* mRNA intact. These results indicate that RNAi ablates target mRNAs through the generation of a sequence-specific nuclease activity. We have termed this enzyme RISC (RNA-induced silencing complex). Although we occasionally observed possible intermediates in the degradation process (see Fig. 2), the absence of stable cleavage end-products indicates an exonuclease (perhaps coupled to an endonuclease). However, it is possible that the RNAi nuclease makes an initial endonucleolytic cut and that non-specific exonucleases in the extract complete the degradation process<sup>16</sup>. In addition, our ability to create an extract that targets *lacZ* *in vitro* indicates that the presence of an endogenous gene is not required for the RNAi response.

To examine the substrate requirements for the dsRNA-induced, sequence-specific nuclease activity, we incubated a variety of *cyclin-E*-derived transcripts with an extract derived from cells that had been transfected with the 540-nucleotide *cyclin E* dsRNA (Fig. 2b, c). Just as a length requirement was observed for the transfected dsRNA, the RNAi nuclease activity showed a dependence on the size of the RNA substrate. Both a 600-nucleotide transcript that extends slightly beyond the targeted region (Fig. 2b) and an ~1-kilobase (kb) transcript that contains the entire coding sequence (data not shown) were completely destroyed by the extract. Surprisingly, shorter substrates were not degraded as

efficiently. Reduced activity was observed against either a 300- or a 220-nucleotide transcript, and a 100-nucleotide transcript was resistant to nuclease in our assay. This was not due solely to position effects because ~100-nucleotide transcripts derived from other portions of the transfected dsRNA behaved similarly (data not shown). As expected, the nuclease activity (or activities) present in the extract could also recognize the antisense strand of the *cyclin E* mRNA. Again, substrates that contained a substantial portion of the targeted region were degraded efficiently whereas those that contained a shorter stretch of homologous sequence (~130 nucleotides) were recognized inefficiently (Fig. 2c, as600). For both the sense and antisense strands, transcripts that had no homology with the transfected dsRNA ( Fig. 2b, Eout; Fig. 2c, as300) were not degraded. Although we cannot exclude the possibility that nuclease specificity could have migrated beyond the targeted region, the resistance of transcripts that do not contain homology to the dsRNA is consistent with data from *C. elegans*. Double-stranded RNAs homologous to an upstream cistron have little or no effect on a linked downstream cistron, despite the fact that unprocessed, polycistronic mRNAs can be readily detected<sup>17, 18</sup>. Furthermore, the nuclease was inactive against a dsRNA identical to that used to provoke the RNAi response *in vivo* (Fig. 2b). In the *in vitro* system, neither a 5' cap nor a poly(A) tail was required, as such transcripts were degraded as efficiently as uncapped and non-polyadenylated RNAs.

Gene silencing provoked by dsRNA is sequence specific. A plausible mechanism for determining specificity would be incorporation of nucleic-acid guide sequences into the complexes that accomplish silencing<sup>19</sup>. In accord with this idea, pre-treatment of extracts with a Ca<sup>2+</sup>-dependent nuclease (micrococcal nuclease) abolished the ability of these extracts to degrade cognate mRNAs (Fig. 3). Activity could not be rescued by addition of non-specific RNAs such as yeast transfer RNA. Although micrococcal nuclease can degrade both DNA and RNA, treatment of the extract with DNase I had no effect (Fig. 3). Sequence-specific nuclease activity, however, did require protein (data not shown). Together, our results support the possibility that the RNAi nuclease is a ribonucleoprotein, requiring both RNA and protein components. Biochemical fractionation (see below) is consistent with these components being associated in extract rather than being assembled on the target mRNA after its addition.

In plants, the phenomenon of co-suppression has been associated with the existence of small (~25-nucleotide) RNAs that correspond to the gene that is being silenced<sup>19</sup>. To address the possibility that a similar RNA might exist in *Drosophila* and guide the sequence-specific nuclease in the choice of substrate, we partially purified our activity through several fractionation steps. Crude extracts contained both sequence-specific nuclease activity and abundant, heterogeneous RNAs homologous to the transfected dsRNA (Figs 2 and 4a). The RNAi nuclease fractionated with ribosomes in a high-speed centrifugation step. Activity could be extracted by treatment with high salt, and

ribosomes could be removed by an additional centrifugation step. Chromatography of soluble nuclease over an anion-exchange column resulted in a discrete peak of activity (Fig. 4b, *cyclin E*). This retained specificity as it was inactive against a heterologous mRNA (Fig. 4b, *lacZ*). Active fractions also contained an RNA species of 25 nucleotides that is homologous to the *cyclin E* target (Fig. 4b, northern). The band observed on northern blots may represent a family of discrete RNAs because it could be detected with probes specific for both the sense and antisense *cyclin E* sequences and with probes derived from distinct segments of the dsRNA (data not shown). At present, we cannot determine whether the 25-nucleotide RNA is present in the nuclease complex in a double-stranded or single-stranded form.

RNA interference allows an adaptive defence against both exogenous and endogenous dsRNAs, providing something akin to a dsRNA immune response. Our data, and that of others<sup>19</sup>, is consistent with a model in which dsRNAs present in a cell are converted, either through processing or replication, into small specificity determinants of discrete size in a manner analogous to antigen processing. Our results suggest that the post-transcriptional component of dsRNA-dependent gene silencing is accomplished by a sequence-specific nuclease that incorporates these small RNAs as guides that target specific messages based upon sequence recognition. The identical size of putative specificity determinants in plants<sup>19</sup> and animals predicts a conservation of both the mechanisms and the components of dsRNA-induced, post-transcriptional gene silencing in diverse organisms. In plants, dsRNAs provoke not only post-transcriptional gene silencing but also chromatin remodelling and transcriptional repression<sup>20, 21</sup>. It is now critical to determine whether conservation of gene-silencing mechanisms also exists at the transcriptional level and whether chromatin remodelling can be directed in a sequence-specific fashion by these same dsRNA-derived guide sequences.

## Methods

**Cell culture and RNA methods** S2 (ref. 22) cells were cultured at 27 °C in 90% Schneider's insect media (Sigma), 10% heat inactivated fetal bovine serum (FBS). Cells were transfected with dsRNA and plasmid DNA by calcium phosphate co-precipitation<sup>23</sup>. Identical results were observed when cells were transfected using lipid reagents (for example, Superfect, Qiagen). For FACS analysis, cells were additionally transfected with a vector that directs expression of a green fluorescent protein (GFP)–US9 fusion protein<sup>13</sup>. These cells were fixed in 90% ice-cold ethanol and stained with propidium iodide at 25 µg ml<sup>-1</sup>. FACS was performed on an Elite flow cytometer (Coulter). For northern blotting, equal loading was ensured by over-probing blots with a control complementary DNA (RP49). For the production of dsRNA, transcription templates were generated by



polymerase chain reaction such that they contained T7 promoter sequences on each end of the template. RNA was prepared using the RiboMax kit (Promega). Confirmation that RNAs were double stranded came from their complete sensitivity to RNase III (a gift from A. Nicholson). Target mRNA transcripts were synthesized using the Riboprobe kit  
5 (Promega) and were gel purified before use.

**Extract preparation** Log-phase S2 cells were plated on 15-cm tissue culture dishes and transfected with 30 µg dsRNA and 30 µg carrier plasmid DNA. Seventy-two hours after transfection, cells were harvested in PBS containing 5 mM EGTA washed twice in PBS and once in hypotonic buffer (10 mM HEPES pH 7.3, 6 mM β-mercaptoethanol). Cells  
10 were suspended in 0.7 packed-cell volumes of hypotonic buffer containing *Complete* protease inhibitors (Boehringer) and 0.5 units ml<sup>-1</sup> of RNasin (Promega). Cells were disrupted in a dounce homogenizer with a type B pestle, and lysates were centrifuged at 30,000g for 20 min. Supernatants were used in an *in vitro* assay containing 20 mM HEPES pH 7.3, 110 mM KOAc, 1 mM Mg(OAc)<sub>2</sub>, 3 mM EGTA, 2 mM CaCl<sub>2</sub>, 1 mM DTT.  
15 Typically, 5 µl extract was used in a 10 µl assay that contained also 10,000 c.p.m. synthetic mRNA substrate.

**Extract fractionation** Extracts were centrifuged at 200,000g for 3 h and the resulting pellet (containing ribosomes) was extracted in hypotonic buffer containing also 1 mM MgCl<sub>2</sub> and 300 mM KOAc. The extracted material was spun at 100,000g for 1 h and the  
20 resulting supernatant was fractionated on Source 15Q column (Pharmacia) using a KCl gradient in buffer A (20 mM HEPES pH 7.0, 1 mM dithiothreitol, 1 mM MgCl<sub>2</sub>). Fractions were assayed for nuclease activity as described above. For northern blotting, fractions were proteinase K/SDS treated, phenol extracted, and resolved on 15% acrylamide 8M urea gels. RNA was electroblotted onto Hybond N+ and probed with  
25 strand-specific riboprobes derived from cyclin E mRNA. Hybridization was carried out in 500 mM NaPO<sub>4</sub> pH 7.0, 15% formamide, 7% SDS, 1% BSA. Blots were washed in 1 SSC at 37–45 °C.

#### References cited in Example 1

- 30 1. Sharp, P. A. RNAi and double-strand RNA. *Genes Dev.* **13**, 139-141 (1999).
2. Sanchez-Alvarado, A. & Newmark, P. A. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl Acad. Sci. USA* **96**, 5049-5054 (1999).
3. Lohmann, J. U., Endl, I. & Bosch, T. C. Silencing of developmental genes in Hydra.  
35 *Dev. Biol.* **214**, 211-214 (1999).



17. Tabara, H., Grishok, A. & Mello, C. C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430-432 (1998).
18. Boshier, J. M., Dufourcq, P., Sookhareea, S. & Labouesse, M. RNA interference can target pre-mRNA. Consequences for gene expression in a *Caenorhabditis elegans* operon. *Genetics* **153**, 1245-1256 (1999).
19. Hamilton, J. A. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950-952 (1999).
20. Jones, L. A., Thomas, C. L. & Maule, A. J. De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. *EMBO J.* **17**, 6385-6393 (1998).
21. Jones, L. A. et al. RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**, 2291-2301 (1999).
22. Schneider, I. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morpho.* **27**, 353-365 (1972).
23. Di Nocera, P. P. & Dawid, I. B. Transient expression of genes introduced into cultured cells of *Drosophila*. *Proc. Natl Acad. Sci. USA* **80**, 7095-7098 (1983).

**Example 2: Role for a bidentate ribonuclease in the initiation step of RNA interference**

Genetic approaches in worms, fungi and plants have identified a group of proteins that are essential for double-stranded RNA-induced gene silencing. Among these are ARGONAUTE family members (e.g. RDE1, QDE2)<sup>9,10,30</sup>, recQ-family helicases (MUT-7, QDE3)<sup>11,12</sup>, and RNA-dependent RNA polymerases (e.g. EGO-1, QDE1, SGS2/SDE1)<sup>13-16</sup>. While potential roles have been proposed, none of these genes has been assigned a definitive function in the silencing process. Biochemical studies have suggested that PTGS is accomplished by a multicomponent nuclease that targets mRNAs for degradation<sup>6,8,17</sup>. We have shown that the specificity of this complex may derive from the incorporation of a small guide sequence that is homologous to the mRNA substrate<sup>6</sup>. Originally identified in plants that were actively silencing transgenes<sup>7</sup>, these ~22 nt. RNAs have been produced during RNAi *in vitro* using an extract prepared from *Drosophila* embryos<sup>8</sup>. Putative guide RNAs can also be produced in extracts from *Drosophila* S2 cells (Fig. 5a). With the goal of understanding the mechanism of post-transcriptional gene silencing, we have undertaken both biochemical fractionation and candidate gene approaches to identify the enzymes that execute each step of RNAi.

Our previous studies resulted in the partial purification of a nuclease, RISC, that is an effector of RNA interference. See Example 1. This enzyme was isolated from *Drosophila* S2 cells in which RNAi had been initiated *in vivo* by transfection with dsRNA. We first sought to determine whether the RISC enzyme and the enzyme that initiates RNAi via processing of dsRNA into 22mers are distinct activities. RISC activity could be largely cleared from extracts by high-speed centrifugation (100,000xg for 60 min.) while the activity that produces 22mers remained in the supernatant (Fig. 5b,c). This simple fractionation indicated that RISC and the 22mer-generating activity are separable and thus distinct enzymes. However, it seems likely that they might interact at some point during the silencing process.

RNase III family members are among the few nucleases that show specificity for double-stranded RNA<sup>18</sup>. Analysis of the *Drosophila* and *C. elegans* genomes reveals several types of RNase III enzymes. First is the canonical RNase III which contains a single RNase III signature motif and a double-stranded RNA binding domain (dsRBD; e.g. RNC\_CAEEL). Second is a class represented by Drosha<sup>19</sup>, a *Drosophila* enzyme that contains two RNase III motifs and a dsRBD (CeDrosha in *C. elegans*). A third class contains two RNase III signatures and an amino terminal helicase domain (e.g. *Drosophila* CG4792, CG6493, *C. elegans* K12H4.8), and these had previously been proposed by Bass as candidate RNAi nucleases<sup>20</sup>. Representatives of all three classes were tested for the ability to produce discrete, ~22 nt. RNAs from dsRNA substrates.

Partial digestion of a 500 nt. cyclin E dsRNA with purified, bacterial RNase III produced a smear of products while nearly complete digestion produced a heterogeneous group of ~11-17 nucleotide RNAs (not shown). In order to test the dual-RNase III enzymes, we prepared T7 epitope-tagged versions of Drosha and CG4792. These were expressed in transfected S2 cells and isolated by immunoprecipitation using antibody-agarose conjugates. Treatment of the dsRNA with the CG4792 immunoprecipitate yielded ~22 nt. fragments similar to those produced in either S2 or embryo extracts (Fig. 6a). Neither activity in extract nor activity in immunoprecipitates depended on the sequence of the RNA substrate since dsRNAs derived from several genes were processed equivalently (see Supplement 1). Negative results were obtained with Drosha and with immunoprecipitates of a DEXH box helicase (Homeless<sup>21</sup>; see Fig 6a,b). Western blotting confirmed that each of the tagged proteins was expressed and immunoprecipitated similarly (see Supplement 2). Thus, we conclude that CG4792 may carry out the initiation step of RNA interference by producing ~22 nt. guide sequences from dsRNAs. Because of its ability to digest dsRNA into uniformly sized, small RNAs, we have named this enzyme Dicer (*Dcr*). *Dicer* mRNA is expressed in embryos, in S2 cells, and in adult flies, consistent with the presence of functional RNAi machinery in all of these contexts (see Supplement 3).

The possibility that Dicer might be the nuclease responsible for the production of guide RNAs from dsRNAs prompted us to raise an antiserum directed against the carboxy-terminus of the Dicer protein (Dicer-1, CG4792). This antiserum could immunoprecipitate a nuclease activity from either *Drosophila* embryo extracts or from S2 cell lysates that produced ~22 nt. RNAs from dsRNA substrates (Fig. 6C). The putative guide RNAs that are produced by the Dicer-1 enzyme precisely comigrate with 22mers that are produced in extract and with 22mers that are associated with the RISC enzyme (Fig. 6 D,F). It had previously been shown that the enzyme that produced guide RNAs in *Drosophila* embryo extracts was ATP-dependent<sup>8</sup>. Depletion of this cofactor resulted in an ~6-fold lower rate of dsRNA cleavage and in the production of RNAs with a slightly lower mobility. Of interest was the fact that both Dicer-1 immunoprecipitates and extracts from S2 cells require ATP for the production of ~22mers (Fig. 6D). We do not observe the accumulation of lower mobility products in these cases, although we do routinely observe these in ATP-depleted embryo extracts. The requirement of this nuclease for ATP is a quite unusual property. We hypothesize that this requirement could indicate that the enzyme may act processively on the dsRNA, with the helicase domain harnessing the energy of ATP hydrolysis both for unwinding guide RNAs and for translocation along the substrate.

Efficient induction of RNA interference in *C. elegans* and in *Drosophila* has several requirements. For example, the initiating RNA must be double-stranded, and it must be several hundred nucleotides in length. To determine whether these requirements are dictated by Dicer, we characterized the ability of extracts and of immunoprecipitated enzyme to digest various RNA substrates. Dicer was inactive against single stranded RNAs regardless of length (see Supplement 4). The enzyme could digest both 200 and 500 nucleotide dsRNAs but was significantly less active with shorter substrates (see Supplement 4). Double-stranded RNAs as short as 35 nucleotides could be cut by the enzyme, albeit very inefficiently (data not shown). In contrast, *E. coli* RNase III could digest to completion dsRNAs of 35 or 22 nucleotides (not shown). This suggests that the substrate preferences of the Dicer enzyme may contribute to but not wholly determine the size dependence of RNAi.

To determine whether the Dicer enzyme indeed played a role in RNAi *in vivo*, we sought to deplete Dicer activity from S2 cells and test the effect on dsRNA-induced gene silencing. Transfection of S2 cells with a mixture of dsRNAs homologous to the two *Drosophila* Dicer genes (CG4792 and CG6493) resulted in an ~6-7 fold reduction of Dicer activity either in whole cell lysates or in Dicer-1 immunoprecipitates (Fig. 7A,B). Transfection with a control dsRNA (murine caspase 9) had no effect. Qualitatively similar results were seen if Dicer was examined by Northern blotting (not shown). Depletion of Dicer in this manner substantially compromised the ability of cells to silence subsequently

an exogenous, GFP transgene by RNAi (Fig. 7C). These results indicate that Dicer is involved in RNAi *in vivo*. The lack of complete inhibition of silencing could result from an incomplete suppression of Dicer (which is itself required for RNAi) or could indicate that *in vivo*, guide RNAs can be produced by more than one mechanism (e.g. through the action of RNA-dependent RNA polymerases).

Our results indicate that the process of RNA interference can be divided into at least two distinct steps. According to this model, initiation of PTGS would occur upon processing of a double-stranded RNA by Dicer into ~22 nucleotide guide sequences, although we cannot formally exclude the possibility that another, Dicer-associated nuclease may participate in this process. These guide RNAs would be incorporated into a distinct nuclease complex (RISC) that targets single-stranded mRNAs for degradation. An implication of this model is that guide sequences are themselves derived directly from the dsRNA that triggers the response. In accord with this model, we have demonstrated that <sup>32</sup>P-labeled, exogenous dsRNAs that have been introduced into S2 cells by transfection are incorporated into the RISC enzyme as 22 mers (Fig. 7E). However, we cannot exclude the possibility that RNA-dependent RNA polymerases might amplify 22mers once they have been generated or provide an alternative method for producing guide RNAs.

The structure of the Dicer enzyme provokes speculation on the mechanism by which the enzyme might produce discretely sized fragments irrespective of the sequence of the dsRNA (see Supplement 1, Fig. 8a). It has been established that bacterial RNase III acts on its substrate as a dimer<sup>18,22,23</sup>. Similarly, a dimer of Dicer enzymes may be required for cleavage of dsRNAs into ~22 nt. pieces. According to one model, the cleavage interval would be determined by the physical arrangement of the two RNase III domains within Dicer enzyme (Fig. 8a). A plausible alternative model would dictate that cleavage was directed at a single position by the two RIII domains in a single Dicer protein. The 22 nucleotide interval could be dictated by interaction of neighboring Dicer enzymes or by translocation along the mRNA substrate. The presence of an integral helicase domain suggests that the products of Dicer cleavage might be single-stranded 22 mers that are incorporated into the RISC enzyme as such.

A notable feature of the Dicer family is its evolutionary conservation. Homologs are found in *C. elegans* (K12H4.8), *Arabidopsis* (e.g., CARPEL FACTORY<sup>24</sup>, T25K16.4, AC012328\_1), mammals (Helicase-MOI<sup>25</sup>) and *S. pombe* (YC9A\_SCHPO) (Fig 8b, see Supplements 6,7 for sequence comparisons). In fact, the human Dicer family member is capable of generating ~22 nt. RNAs from dsRNA substrates (Supplement 5) suggesting that these structurally similar proteins may all share similar biochemical functions. It has been demonstrated that exogenous dsRNAs can affect gene function in early mouse

embryos<sup>29</sup>, and our results suggest that this regulation may be accomplished by an evolutionarily conserved RNAi machinery.

In addition to RNaseIII and helicase motifs, searches of the PFAM database indicate that each Dicer family member also contains a ZAP domain (Fig 8c)<sup>27</sup>. This sequence was defined based solely upon its conservation in the Zwille/ARGONAUTE/Piwi family that has been implicated in RNAi by mutations in *C. elegans* (*Rde-1*)<sup>9</sup> and *Neurospora* (*Qde-2*)<sup>10</sup>. Although the function of this domain is unknown, it is intriguing that this region of homology is restricted to two gene families that participate in dsRNA-dependent silencing. Both the ARGONAUTE and Dicer families have also been implicated in common biological processes, namely the determination of stem-cell fates. A hypomorphic allele of *carpel factory*, a member of the Dicer family in *Arabidopsis*, is characterized by increased proliferation in floral meristems<sup>24</sup>. This phenotype and a number of other characteristic features are also shared by *Arabidopsis* ARGONAUTE (*ago1-1*) mutants<sup>26</sup> (C. Kidner and R. Martienssen, pers. comm.). These genetic analyses begin to provide evidence that RNAi may be more than a defensive response to unusual RNAs but may also play important roles in the regulation of endogenous genes.

With the identification of Dicer as a catalyst of the initiation step of RNAi, we have begun to unravel the biochemical basis of this unusual mechanism of gene regulation. It will be of critical importance to determine whether the conserved family members from other organisms, particularly mammals, also play a role in dsRNA-mediated gene regulation.

## Methods

Plasmid constructs. A full-length cDNA encoding Droscha was obtained by PCR from an EST sequenced by the Berkeley Drosophila genome project. The *Homeless* clone was a gift from Gillespie and Berg (Univ. Washington). The T7 epitope-tag was added to the amino terminus of each by PCR, and the tagged cDNAs were cloned into pRIP, a retroviral vector designed specifically for expression in insect cells (E. Bernstein, unpublished). In this vector, expression is driven by the *Orgyia pseudotsugata* IE2 promoter (Invitrogen). Since no cDNA was available for CG4792/Dicer, a genomic clone was amplified from a bacmid (BACR23F10; obtained from the BACPAC Resource Center in the Dept. of Human Genetics at the Roswell Park Cancer Institute). Again, during amplification, a T7 epitope tag was added at the amino terminus of the coding sequence. The human Dicer gene was isolated from a cDNA library prepared from HaCaT cells (GJH, unpublished). A T7-tagged version of the complete coding sequence was cloned into pCDNA3 (Invitrogen) for expression in human cells (LinX-A).

Cell culture and extract preparation. *S2 and embryo culture.* S2 cells were cultured at 27°C in 5% CO<sub>2</sub> in Schneider's insect media supplemented with 10% heat inactivated fetal bovine serum (Gemini) and 1% antibiotic-antimycotic solution (Gibco BRL). Cells were harvested for extract preparation at 10x10<sup>6</sup> cells/ml. The cells were washed 1X in PBS and were resuspended in a hypotonic buffer (10 mM Hepes pH 7.0, 2mM MgCl<sub>2</sub>, 6 mM βME) and dounced. Cell lysates were spun 20,000xg for 20 minutes. Extracts were stored at -80°C. *Drosophila* embryos were reared in fly cages by standard methodologies and were collected every 12 hours. The embryos were dechorionated in 50% chlorox bleach and washed thoroughly with distilled water. Lysis buffer (10mM Hepes, 10mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.5mM EGTA, 10mM β-glycerophosphate, 1mM DTT, 0.2 mM PMSF) was added to the embryos, and extracts were prepared by homogenization in a tissue grinder. Lysates were spun for two hours at 200,000xg and were frozen at -80°C. LinX-A cells, a highly-transfectable derivative of human 293 cells, (Lin Xie and GJH, unpublished) were maintained in DMEM/10%FCS.

Transfections and immunoprecipitations. S2 cells were transfected using a calcium phosphate procedure essentially as previously described<sup>6</sup>. Transfection rates were ~90% as monitored in controls using an *in situ* β-galactosidase assay. LinX-A cells were also transfected by calcium phosphate co-precipitation. For immunoprecipitations, cells (~5x10<sup>6</sup> per IP) were transfected with various clones and lysed three days later in IP buffer (125mM KOAc, 1mM MgOAc, 1mM CaCl<sub>2</sub>, 5mM EGTA, 20mM Hepes pH 7.0, 1mM DTT, 1% NP-40 plus Complete protease inhibitors (Roche)). Lysates were spun for 10 minutes at 14,000xg and supernatants were added to T7 antibody-agarose beads (Novagen). Antibody binding proceeded for 4 hours at 4°C. Beads were centrifuged and washed in lysis buffer three times, and once in reaction buffer. The Dicer antiserum was raised in rabbits using a KLH-conjugated peptide corresponding to the C-terminal 8 amino acids of *Drosophila* Dicer-1 (CG4792).

Cleavage reactions. *RNA preparation.* Templates to be transcribed into dsRNA were generated by PCR with forward and reverse primers, each containing a T7 promoter sequence. RNAs were produced using Riboprobe (Promega) kits and were uniformly labeling during the transcription reaction with <sup>32</sup>P-UTP. Single-stranded RNAs were purified from 1% agarose gels. *dsRNA cleavage.* Five microliters of embryo or S2 extracts were incubated for one hour at 30°C with dsRNA in a reaction containing 20mM Hepes pH 7.0, 2mM MgOAc, 2mM DTT, 1mM ATP and 5% Superasin (Ambion). Immunoprecipitates were treated similarly except that a minimal volume of reaction buffer (including ATP and Superasin) and dsRNA were added to beads that had been washed in reaction buffer (see above). For ATP depletion, *Drosophila* embryo extracts were incubated for 20 minutes at 30°C with 2mM glucose and 0.375 U of hexokinase (Roche) prior to the addition of dsRNA.



Northern and Western analysis. Total RNA was prepared from *Drosophila* embryos (0-12 hour), from adult flies, and from S2 cells using Trizol (Lifetech). Messenger RNA was isolated by affinity selection using magnetic oligo-dT beads (Dynal). RNAs were electrophoresed on denaturing formaldehyde/agarose gels, blotted and probed with randomly primed DNAs corresponding to Dicer. For Western analysis, T7-tagged proteins were immunoprecipitated from whole cell lysates in IP buffer using anti-T7-antibody-agarose conjugates. Proteins were released from the beads by boiling in Laemmli buffer and were separated by electrophoresis on 8% SDS PAGE. Following transfer to nitrocellulose, proteins were visualized using an HRP-conjugated anti-T7 antibody (Novagen) and chemiluminescent detection (Supersignal, Pierce).

RNAi of Dicer. *Drosophila* S2 cells were transfected either with a dsRNA corresponding to mouse caspase 9 or with a mixture of two dsRNAs corresponding to *Drosophila* Dicer-1 and Dicer-2 (CG4792 and CG6493). Two days after the initial transfection, cells were again transfected with a mixture containing a GFP expression plasmid and either luciferase dsRNA or GFP dsRNA as previously described<sup>6</sup>. Cells were assayed for Dicer activity or fluorescence three days after the second transfection. Quantification of fluorescent cells was done on a Coulter EPICS cell sorter after fixation. Control transfections indicated that Dicer activity was not affected by the introduction of caspase 9 dsRNA.

#### References cited Example 2

1. Baulcombe, D. C. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol Biol* 32, 79-88 (1996).
2. Wassenegger, M. & Pelissier, T. A model for RNA-mediated gene silencing in higher plants. *Plant Mol Biol* 37, 349-62 (1998).
3. Montgomery, M. K. & Fire, A. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression [see comments]. *Trends Genet* 14, 255-8 (1998).
4. Sharp, P. A. RNAi and double-strand RNA. *Genes Dev* 13, 139-41 (1999).
5. Sijen, T. & Kooter, J. M. Post-transcriptional gene-silencing: RNAs on the attack or on the defense? [In Process Citation]. *Bioessays* 22, 520-31 (2000).
6. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293-6 (2000).
7. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants [see comments]. *Science* 286, 950-2 (1999).



22. Robertson, H. D., Webster, R. E. & Zinder, N. D. Purification and properties of ribonuclease III from *Escherichia coli*. *J Biol Chem* 243, 82-91 (1968).
23. Dunn, J. J. RNase III cleavage of single-stranded RNA. Effect of ionic strength on the fidelity of cleavage. *J Biol Chem* 251, 3807-14 (1976).
- 5 24. Jacobsen, S. E., Running, M. P. & Meyerowitz, E. M. Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* 126, 5231-43 (1999).
25. Matsuda, S. *et al.* Molecular cloning and characterization of a novel human gene (HERNA) which encodes a putative RNA-helicase. *Biochim Biophys Acta* 1490, 163-9  
10 (2000).
26. Bohmert, K. *et al.* AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. *Embo J* 17, 170-80 (1998).
27. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28, 405-20 (1997).
- 15 28. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402 (1997).
29. Wianny, F. and Zernicka-Goetz, M. Specific interference with gene function by double-stranded RNA in early mouse development. *Nature Cell Biol.* 2, 70-75 (2000).
30. Fagard, M., Boutet, S., Morel, J.-B., Bellini, C. and Vaucheret, H. Ago-1, Qde-2  
20 and Rde-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl. Acad. Sci. USA* 97, 11650-11654 (2000).

25 **Example 3: A simplified method for the creation of hairpin constructs for RNA interference.**

In numerous model organisms, double stranded RNAs have been shown to cause effective and specific suppression of gene function (ref. 1). This response, termed RNA interference or post-transcriptional gene silencing, has evolved into a highly effective reverse genetic tool in *C. elegans*, *Drosophila*, plants and numerous other systems. In  
30 these cases, double-stranded RNAs can be introduced by injection, transfection or feeding; however, in all cases, the response is both transient and systemic. Recently, stable interference with gene expression has been achieved by expression of RNAs that form snap-back or hairpin structures (refs 2-7). This has the potential not only to allow stable silencing of gene expression but also inducible silencing as has been observed in  
35 trypanosomes and adult *Drosophila* (refs 2,4,5). The utility of this approach is somewhat

hampered by the difficulties that arise in the construction of bacterial plasmids containing the long inverted repeats that are necessary to provoke silencing. In a recent report, it was stated that more than 1,000 putative clones were screened to identify the desired construct (ref 7).

5           The presence of hairpin structures often induces plasmid rearrangement, in part due to the E. coli sbc proteins that recognize and cleave cruciform DNA structures (ref 8). We have developed a method for the construction of hairpins that does not require cloning of inverted repeats, per se. Instead, the fragment of the gene that is to be silenced is cloned as a direct repeat, and the inversion is accomplished by treatment with a site-  
10 specific recombinase, either *in vitro* (or potentially *in vivo*) (see Fig 27). Following recombination, the inverted repeat structure is stable in a bacterial strain that lacks an intact SBC system (DL759). We have successfully used this strategy to construct numerous hairpin expression constructs that have been successfully used to provoke gene silencing in *Drosophila* cells.

15

#### Literature Cited in Example 3

1. Boshier, J. M. & Labouesse, M. RNA interference: genetic wand and genetic watchdog. *Nat Cell Biol* 2, E31-6 (2000).
2. Fortier, E. & Belote, J. M. Temperature-dependent gene silencing by an expressed inverted repeat in *Drosophila* [published erratum appears in *Genesis*;2000  
20 May;27(1):47]. *Genesis* 26, 240-4 (2000).
3. Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat Biotechnol* 18, 896-8 (2000).
4. Lam, G. & Thummel, C. S. Inducible expression of double-stranded RNA directs  
25 specific genetic interference in *Drosophila* [In Process Citation]. *Curr Biol* 10, 957-63 (2000).
5. Shi, H. *et al.* Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *Rna* 6, 1069-76 (2000).
6. Smith, N. A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* 407,  
30 319-20 (2000).
7. Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A. & Driscoll, M. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nat Genet* 24, 180-3 (2000).

